การเปรียบเทียบประสิทธิภาพการทดสอบความเท่ากันของค่ากลาง เมื่อข้อมูลอยู่ในมาตราอันดับและมีความสัมพันธ์กัน

A Comparison on Performance of Tests for Equality of Means of Related Ordinal Data

วนิดา พงษ์ศักดิ์ชาติ* และ นพรัตน์ แป้นงาม

Vanida Pongsakchat* and Nopparat Panngam

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

Department of Mathematics, Faculty of Science, Burapha University

Received : 17 May 2018

Accepted : 26 July 2018

Published online : 1 August 2018

บทคัดย่อ

ในการทดสอบสมมุติฐานเกี่ยวกับค่ากลางของประชากรมากกว่า 1 กลุ่ม เมื่อข้อมูลแต่ละกลุ่มมีความสัมพันธ์กัน การทดสอบเชิงสถิติที่ใช้มักเป็นการทดสอบแบบอิงพารามิเตอร์ที่มีข้อสมมุติว่าข้อมูลที่ศึกษาต้องเป็นข้อมูลเชิงปริมาณ และมีการแจกแจงปรกติ อย่างไรก็ตามหากข้อสมมุติเหล่านี้ไม่เป็นไปตามที่กำหนด นั่นคือข้อมูลที่ศึกษาอยู่ในมาตราอันดับแบบมาตรวัดของลิเกิร์ต 5 ระดับ และ/หรือ ไม่มีการแจกแจงปรกติ การทดสอบเชิงสถิติที่เป็นทางเลือกคือการทดสอบไม่อิงพารามิเตอร์ ในงานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพของการทดสอบทีและการทดสอบวิลค็อกซัน สำหรับการทดสอบค่ากลางของประชากร 2 กลุ่ม และการทดสอบเอฟและการทดสอบ Friedman สำหรับการทดสอบค่ากลางของประชากร 3 กลุ่ม เมื่อประชากรของข้อมูลมีการแจกแจงชนิดต่าง ๆ 4 ชนิด ได้แก่ การแจกแจงปรกติ การแจกแจงเอกรูป การแจกแจงเบ้ซ้าย และการแจกแจงเบ้ขวา และมีขนาดของความสัมพันธ์เท่ากับ 0.50 และ 0.70 กำหนดขนาดตัวอย่างเท่ากับ 10, 20, 30, 50 และ 100 ในกรณีของการทดสอบ 2 กลุ่ม และกำหนดขนาดตัวอย่างเท่ากับ 20, 30, 50 และ 100 ในกรณีของการทดสอบ 3 กลุ่ม จากการศึกษาพบว่า เมื่อตัวอย่างมีขนาดเล็ก (10, 20 และ 30 สำหรับการทดสอบ 2 กลุ่ม, 20 และ 30 สำหรับการทดสอบ 3 กลุ่ม) การทดสอบแบบอิงพารามิเตอร์ คือ การทดสอบทีและการทดสอบเอฟมีประสิทธิภาพดีกว่าการทดสอบไม่อิงพารามิเตอร์ คือ วิลค็อกซันและการทดสอบ Friedman ทั้งในส่วนของความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ และเมื่อตัวอย่างมีขนาดใหญ่ขึ้นการทดสอบทั้งสองชนิดมีประสิทธิภาพใกล้เคียงกัน โดยการทดสอบทั้งสองชนิดมีกำลังการทดสอบสูงขึ้นเมื่อขนาดตัวอย่างและสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้น และการแจกแจงของประชากรที่ศึกษาในแต่ละกลุ่มมีความแตกต่างกันมาก

คำสำคัญ : การทดสอบที, การทดสอบเอฟ, การทดสอบวิลค็อกซัน, การทดสอบ Friedman

*Corresponding author. E-mail : vanida@buu.ac.th

Abstract

For hypothesis testing about more than one population mean when the data are related, parametric tests usually have been used. When using parametric tests, assumptions are required, the data have to be quantitative data and normally distributed. However, if the data are ordinal Likert scale and/or non-normally distributed, non-parametric tests are alternative. In this research, in case of testing two population means, *t*-test and Wilcoxon test are compared and for testing three population means, *F*-test and Friedman test are compared. The correlated ordinal data are generated from normal, uniform, left-skewed and right-skewed distributions. In addition, the correlation coefficients are 0.50 and 0.70. For testing two population means, sample sizes are 10, 20, 30, 50 and 100 and for testing three population means, sample sizes are 20, 30, 50 and 100. From the simulation study, when the sample sizes are small (10, 20 and 30 for two groups test, 20 and 30 for three groups test), parametric tests (*t*-test and *F*-test) perform better than non-parametric tests (Wilcoxon test and Friedman test) in term of controlling the type I error rate and power. When the sample sizes are medium and large, both methods have similar performances. The power of both methods are increased as sample sizes and correlation increase and the distributions are completely difference.

Keywords : *t*-test, *F*-test, Wilcoxon test, Friedman test

บทนำ

ในงานวิจัยด้านต่าง ๆ จำเป็นต้องใช้วิธีการทางสถิติในการวิเคราะห์ข้อมูลที่เก็บรวบรวมมาได้ เพื่อหาข้อสรุปที่มีความถูกต้องและน่าเชื่อถือ ซึ่งความถูกต้องและน่าเชื่อถือของข้อสรุปที่ได้จะมีมากหรือน้อยนั้นขึ้นอยู่กับองค์ประกอบดังต่อไปนี้ คือ การวางแผนหรือการออกแบบการทดลอง เทคนิคการสุ่มตัวอย่าง และการใช้วิธีเชิงสถิติในการวิเคราะห์ข้อมูลที่ได้มา ซึ่งวิธีเชิงสถิติทุกวิธีต่างมีข้อสมมุติในการนำไปใช้ทั้งสิ้น โดยข้อสมมุติเหล่านี้มักเกี่ยวข้องกับชนิดของข้อมูลและลักษณะการแจกแจงความน่าจะเป็นของข้อมูล หากผู้วิจัยนำวิธีเชิงสถิติไปใช้อย่างไม่ถูกต้อง ไม่สอดคล้องกับข้อสมมุติที่กำหนดไว้อาจส่งผลให้ข้อสรุปของงานวิจัยที่ได้เป็นข้อสรุปที่ไม่ถูกต้อง และก่อให้เกิดผลเสียตามมาได้

ในการศึกษาบางการศึกษาข้อมูลที่เก็บรวบรวมได้ไม่ได้เป็นข้อมูลเชิงปริมาณ แต่เป็นข้อมูลแบบมาตรวัดของลิเกิร์ต ซึ่งอยู่ในมาตราอันดับ (Ordinal scale) เช่น ในงานวิจัยทางด้านการแพทย์เกี่ยวกับผลการรักษาอาการปวดศีรษะไมเกรนโดยใช้ยาชนิดหนึ่ง โดยแพทย์ได้ให้อาสาสมัครที่เข้าร่วมการทดลองได้ประเมินระดับความเจ็บปวดของตนเองในการวัดซ้ำแต่ละครั้งหลังการได้รับยาเป็น ปวดมากที่สุด ปวดมาก ปวดปานกลาง ปวดน้อย ไม่มีอาการปวดเลย หรือในงานวิจัยทางด้านวิทยาศาสตร์การอาหารที่ศึกษาเกี่ยวกับความพึงพอใจของรสชาติของอาหารที่ผลิตด้วยส่วนผสมที่แตกต่างกันโดยใช้ผู้ชิมแต่ละคนได้ประเมินรสชาติของอาหารแต่ละสูตรในลักษณะของระดับความพึงพอใจ คือ พึงพอใจมากที่สุด พึงพอใจมาก พึงพอใจปานกลาง พึงพอใจน้อย และไม่พึงพอใจ ซึ่งข้อมูลที่ได้ในลักษณะนี้ในทางสถิติถือเป็นข้อมูลในมาตราอันดับซึ่งไม่ได้มีค่าเป็นตัวเลขอย่างแท้จริงจึงไม่ใช่ข้อมูลเชิงปริมาณ และไม่ได้มีการแจกแจงปรกติ จึงไม่เป็นไปตามข้อสมมุติของการทดสอบอิงพารามิเตอร์ (Parametric test) การทดสอบเชิงสถิติที่เหมาะสมกับข้อมูลชนิดนี้มากกว่าคือการทดสอบไม่อิงพารามิเตอร์

(Non-parametric test) ได้แก่ การทดสอบวิลค็อกซัน (Wilcoxon test) ในกรณีที่ $K = 2$ และการทดสอบ Friedman (Friedman test) ในกรณีที่ $K = 3, 4, \ldots$ แต่จากการศึกษาค้นคว้าของผู้วิจัยพบว่าในงานวิจัยที่มีข้อมูลลักษณะนี้นักวิจัยมักใช้การทดสอบที (*t*-test) หรือการวิเคราะห์ความแปรปรวนในการวิเคราะห์ข้อมูล ซึ่งเป็นวิธีที่ไม่สอดคล้องกับลักษณะของข้อมูล

จากงานวิจัยที่ผ่านมาได้มีการศึกษาเปรียบเทียบประสิทธิภาพของการทดสอบอิงพารามิเตอร์และไม่อิงพารามิเตอร์สำหรับการทดสอบค่ากลางของประชากร 2 กลุ่ม และหลายกลุ่มเมื่อข้อมูลมีความสัมพันธ์กัน โดย Zimmerman and Zumbo (1993) ได้เปรียบเทียบการทดสอบที การทดสอบวิลค็อกซัน และการทดสอบเครื่องหมาย (Sign test) สำหรับข้อมูล 2 กลุ่ม ที่มีความสัมพันธ์กัน และเปรียบเทียบการทดสอบเอฟ (*F*-test) สำหรับข้อมูลจากการทดลองแบบวัดซ้ำ (Repeated measures designs) การทดสอบเอฟ (*F*-test) สำหรับข้อมูลจากการทดลองแบบวัดซ้ำที่ใช้ลำดับของข้อมูลแทนข้อมูลจริง และการทดสอบ Friedman สำหรับข้อมูล K กลุ่ม เมื่อ K =2, 3 และ 4 โดยข้อมูลที่ศึกษาถูกสุ่มมาจากประชากรที่มีแจกแจงปรกติ การแจกแจงเอกรูป การแจกแจงปรกติแบบผสม การแจกแจงแบบเลขชี้กำลัง การแจกแจงลาปาส และการแจกแจงโคชี ซึ่งเป็นการแจกแจงความน่าจะเป็นของตัวแปรเชิงปริมาณ โดยพบว่าการทดสอบ Friedman มีประสิทธิภาพใกล้เคียงกับการทดสอบเครื่องหมายสำหรับทุกการแจกแจง ในขณะที่การทดสอบเอฟใช้ลำดับของข้อมูลแทนข้อมูลจริงมีประสิทธิภาพใกล้เคียงกับการทดสอบวิลค็อกซัน ต่อมา Harwell and Serlin (1994) ได้ศึกษาเปรียบเทียบการทดสอบเอฟ การทดสอบ Friedman การทดสอบ Aligned rank การทดสอบ Rank transform การทดสอบ Puri and Sen เมื่อข้อมูลมาจากการทดลองแบบวัดซ้ำ ผลการศึกษาพบว่า การทดสอบ Friedman และการทดสอบไม่อิงพารามิเตอร์อื่น ๆ เป็นการทดสอบที่ไม่มีประสิทธิภาพเมื่อประชากรมีการแจกแจงแบบสมมาตร และความแปรปรวนแตกต่างกันมาก และเมื่อประชากรมีการแจกแจงแบบเบ้ และมีความแปรปรวนแตกต่างกันเล็กน้อย อย่างไรก็ตาม การทดสอบเอฟมีความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ดีกว่าการทดสอบไม่อิงพารามิเตอร์

ในกรณีที่ข้อมูลอยู่ในมาตราอันดับ Romano, Kromrey, Coraggio and Skowronek (2006) ได้ศึกษาถึงการทดสอบความแตกต่างกันของข้อมูล 2 กลุ่ม ที่อยู่ในมาตราอันดับ การทดสอบที่ศึกษาได้แก่ การทดสอบที และสถิติ Cohen's *d* ซึ่งถือว่าเป็นการทดสอบอิงพารามิเตอร์ การทดสอบไคกำลังสอง Odds ratios และ Cliff's delta ซึ่งถือว่าเป็นการทดสอบไม่อิงพารามิเตอร์ ผลการศึกษาพบว่าการทดสอบที และ Cohen's *d* เป็นวิธีที่เหมาะสมในกรณีที่ต้องการทดสอบความแตกต่างระหว่างกลุ่มของข้อมูลที่อยู่ในมาตราอันดับ และNorman (2010) ได้ศึกษางานวิจัยย้อนหลังตั้งแต่ปี ค.ศ 1930 ที่แสดงให้เห็นว่าวิธีเชิงสถิติแบบอิงพารามิเตอร์ ได้แก่ การวิเคราะห์ความแปรปรวน การวิเคราะห์การถดถอยและสหสัมพันธ์ เป็นวิธีที่มีความแกร่ง ถึงแม้ว่าข้อมูลที่ศึกษาอยู่ในมาตราอันดับแบบมาตรวัดของลิเกิร์ต ซึ่งไม่เป็นไปตามข้อสมมุติของวิธีเชิงสถิติเหล่านี้ นั่นคือการใช้วิธีเชิงสถิติเหล่านี้ในการวิเคราะห์ข้อมูลยังคงให้ข้อสรุปของงานวิจัยที่ถูกต้อง

จากการที่ข้อมูลมาตราอันดับถูกนำมาใช้ในงานวิจัยสาขาต่าง ๆ และการทดสอบเชิงสถิติที่ใช้สำหรับข้อมูลประเภทนี้มักเป็นการทดสอบอิงพารามิเตอร์ ซึ่งไม่เป็นไปตามข้อสมมุติของการทดสอบ ดังนั้นในงานวิจัยนี้ ผู้วิจัยจึงมีแนวคิดที่จะศึกษาเปรียบเทียบประสิทธิภาพของการทดสอบทั้ง 2 ประเภท ที่ใช้สำหรับการทดสอบความเท่ากันของค่ากลางของประชากรเมื่อข้อมูลอยู่ในมาตราอันดับแบบมาตรวัดของลิเกิร์ต 5 ระดับ และมีความสัมพันธ์ ในกรณีที่ $K = 2$ วิธีการทดสอบที่ศึกษาเปรียบเทียบ คือ การทดสอบทีสำหรับข้อมูลที่ไม่เป็นอิสระกัน และการทดสอบวิลค็อกซัน ในกรณีที่ $K = 3$ วิธีการทดสอบที่ศึกษาเปรียบเทียบ คือ การทดสอบเอฟ และการทดสอบ Friedman เพื่อให้เห็นถึงประสิทธิภาพของการทดสอบเหล่านี้ในกรณีต่าง ๆ

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการศึกษาโดยใช้วิธีมอนติคาร์โล (Monte Carlo method) เพื่อเปรียบเทียบประสิทธิภาพของการทดสอบเชิงสถิติแบบต่าง ๆ โดยพิจารณาจาก 1) ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 (Type I error) และ 2) กำลังการทดสอบ (Power of a test) เมื่อข้อมูลทุกกลุ่มอยู่ในมาตราอันดับแบบมาตรวัดของลิเกิร์ต 5 ระดับ และมีความสัมพันธ์กัน จำนวน 2 และ 3 กลุ่ม โดยมีขั้นตอนการศึกษา ดังนี้

1. สร้างข้อมูลจำนวน K กลุ่ม เมื่อ $K = 2$ และ 3 ด้วยโปรแกรม R ด้วยชุดคำสั่งสำเร็จรูป “GenOrd” ที่พัฒนาโดย Barbiero and Ferrari (2015) ซึ่งเป็นชุดคำสั่งสำหรับสร้างข้อมูลมาตราอันดับ ที่ใช้วิธีการของ Ferrari and Barbiero (2012) ซึ่งเป็นวิธีที่ต้องมีการกำหนดเมทริกซ์สหสัมพันธ์ (Correlation matrix) และการแจกแจงตามขอบ (Marginal distributions) ของข้อมูลไว้ก่อน ให้มีลักษณะการแจกแจงต่าง ๆ 4 ชนิด ได้แก่ การแจกแจงปรกติ การแจกแจงเอกรูป การแจกแจงแบบเบ้ขวา และการแจกแจงแบบเบ้ซ้าย โดยชุดข้อมูลมีความสัมพันธ์กันขนาด 0.5 และ 0.7 และกำหนดตัวอย่างเท่ากับ 10, 20, 30, 50 และ 100 สำหรับการทดสอบประชากร 2 กลุ่ม ($K = 2$) และเท่ากับ 20, 30, 50 และ 100 สำหรับการทดสอบประชากร 3 กลุ่ม ($K = 3$) โดยการแจกแจงความน่าจะเป็นแต่ละแบบมีการกำหนดความน่าจะเป็นที่ข้อมูลจะอยู่ในแต่ละระดับดังที่แสดงในตารางที่ 1

ตารางที่ 1 ความน่าจะเป็นที่ข้อมูลจะอยู่ในแต่ละระดับ (5 ระดับ)

การแจกแจง	ความน่าจะเป็นที่ข้อมูลจะอยู่ในแต่ระดับ				
	1	2	3	4	5
การแจกแจงปรกติ	0.15	0.20	0.30	0.20	0.15
การแจกแจงเอกรูป	0.20	0.20	0.20	0.20	0.20
การแจกแจงเบ้ซ้าย	0.05	0.10	0.20	0.35	0.30
การแจกแจงเบ้ขวา	0.30	0.35	0.20	0.10	0.05

ในกรณีที่ข้อมูลทุกกลุ่มมีการแจกแจงความน่าจะเป็นแบบเดียวกันจะพิจารณาความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกรณีที่ข้อมูล K กลุ่ม มีการแจกแจงความน่าจะเป็นแตกต่างกันอย่างน้อยหนึ่งกลุ่มจะพิจารณากำลังการทดสอบของการทดสอบชนิดต่าง ๆ

2. นำข้อมูลที่ได้ในแต่ละกรณีมาทดสอบสมมุติฐาน โดยมีสมมุติฐานทางสถิติ คือ

H_0 : ประชากร K กลุ่ม มีค่ากลางไม่แตกต่างกัน

H_a : มีประชากรอย่างน้อย 2 กลุ่ม มีค่ากลางแตกต่างกัน

ในกรณีที่ $K = 2$ การทดสอบที่ใช้คือการทดสอบทีสำหรับข้อมูลที่ไม่เป็นอิสระกัน และการทดสอบวิลค็อกซัน

ในกรณีที่ $K = 3$ การทดสอบที่ใช้คือการทดสอบเอฟ และการทดสอบ Friedman

โดยในการทดสอบสมมุติฐานจะใช้ระดับนัยสำคัญเท่ากับ 0.05 และมีเกณฑ์การตัดสินใจปฏิเสธ H_0 เมื่อ *p*-value ของค่าสถิติทดสอบน้อยกว่าระดับนัยสำคัญที่กำหนด

3. ทำซ้ำขั้นตอนที่ 1 และ 2 ในแต่ละกรณีจำนวน 5,000 รอบ แล้วนับจำนวนครั้งของการปฏิเสธ H_0

4. ในกรณีที่ข้อมูลทุกกลุ่มมีการแจกแจงความน่าจะเป็นแบบเดียวกันนั่นคือมีค่ากลางเท่ากัน คำนวณค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 จาก

$$\frac{\text{จำนวนครั้งของการปฏิเสธ } H_0 \text{ เมื่อ } H_0 \text{ จริง}}{5{,}000}$$

และเปรียบเทียบค่าประมาณความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley (1978) ที่ระดับนัยสำคัญ 0.05 โดยช่วงที่กำหนดคือ [0.025,0.075]

5. ในกรณีที่ข้อมูล K กลุ่ม มีการแจกแจงความน่าจะเป็นแตกต่างกันอย่างน้อย 1 กลุ่ม นั่นคือข้อมูลมีค่ากลางแตกต่างกันอย่างน้อย 1 กลุ่ม คำนวณค่ากำลังการทดสอบจาก

$$\frac{\text{จำนวนครั้งของการปฏิเสธ } H_0 \text{ เมื่อ } H_0 \text{ เป็นเท็จ}}{5{,}000}$$

6. ในการพิจารณาประสิทธิภาพของการทดสอบแต่ละวิธี จะพิจารณาจากความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ก่อน แล้วจึงพิจารณากำลังการทดสอบเฉพาะการทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้เท่านั้น โดยการทดสอบที่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ตามเกณฑ์ของ Bradley ได้และมีกำลังการทดสอบสูงที่สุดถือว่าเป็นการทดสอบที่มีประสิทธิภาพมากที่สุด

ผลการวิจัยและวิจารณ์ผล

ผลการวิจัยนี้แบ่งเป็น 2 ส่วน คือความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ

1. ความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1

1.1 เมื่อ K=2 (ผลการศึกษาแสดงในตารางที่ 2)

เมื่อพิจารณาตามเกณฑ์ของ Bradley พบว่าการทดสอบทีสามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในทุกกรณี ในขณะที่การทดสอบวิลคอกซันไม่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ เมื่อขนาดตัวอย่างเท่ากับ 10 แต่สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในกรณีอื่น ๆ

1.2 เมื่อ $K = 3$ (ผลการศึกษาแสดงในตารางที่ 3)

เมื่อพิจารณาตามเกณฑ์ของ Bradley พบว่าการทดสอบเอฟและการทดสอบ Friedman สามารถควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ได้ในทุกกรณี

ตารางที่ 2 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อ $K = 2$

		สัมประสิทธิ์สหสัมพันธ์			
		0.5		0.7	
การแจกแจง	ขนาดตัวอย่าง	การทดสอบที	การทดสอบวิลค็อกซัน	การทดสอบที	การทดสอบวิลค็อกซัน
ปรกติ	10	0.0316	0.0200*	0.0304	0.0170*
	20	0.0468	0.0386	0.0420	0.0316
	30	0.0450	0.0428	0.0490	0.0428
	50	0.0508	0.0492	0.0488	0.0492
	100	0.0512	0.0488	0.0448	0.0450
เอกรูป	10	0.0330	0.0218*	0.0386	0.0190*
	20	0.0404	0.0340	0.0440	0.0342
	30	0.0436	0.0412	0.0484	0.0430
	50	0.0462	0.0440	0.0512	0.0484
	100	0.0552	0.0534	0.0582	0.0548
เบ้ซ้าย	10	0.0322	0.0182*	0.0282	0.0164*
	20	0.0429	0.0356	0.0402	0.0280
	30	0.0439	0.0388	0.0546	0.0474
	50	0.0508	0.0466	0.0516	0.0506
	100	0.0486	0.0478	0.0510	0.0502
เบ้ขวา	10	0.0284	0.0162*	0.0298	0.0160*
	20	0.0414	0.0342	0.0494	0.0386
	30	0.0486	0.0432	0.0470	0.0428
	50	0.0514	0.0484	0.0454	0.0444
	100	0.0504	0.0498	0.0478	0.0458

* ควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 ไม่ได้

ตารางที่ 3 ความน่าจะเป็นของความผิดพลาดแบบที่ 1 เมื่อ $K = 3$

		สัมประสิทธิ์สหสัมพันธ์			
		0.5		0.7	
การแจกแจง	ขนาดตัวอย่าง	การทดสอบเอฟ	การทดสอบ Friedman	การทดสอบเอฟ	การทดสอบ Friedman
ปรกติ	20	0.0468	0.0438	0.0492	0.0450
	30	0.0482	0.0508	0.0474	0.0466
	50	0.0504	0.0450	0.0524	0.0482
	100	0.0484	0.0484	0.0504	0.0494
เอกรูป	20	0.0510	0.0446	0.0510	0.0442
	30	0.0482	0.0456	0.0504	0.0458
	50	0.0486	0.0486	0.0506	0.0484
	100	0.0514	0.0516	0.0554	0.0498
เบ้ซ้าย	20	0.0506	0.0468	0.0446	0.0404
	30	0.0446	0.0412	0.0498	0.0492
	50	0.0506	0.0476	0.0466	0.0442
	100	0.0482	0.0494	0.0492	0.0490
เบ้ขวา	20	0.0466	0.0396	0.0450	0.0390
	30	0.0502	0.0444	0.0468	0.0466
	50	0.0486	0.0458	0.0484	0.0440
	100	0.0492	0.0460	0.0542	0.0540

2. กำลังการทดสอบ

2.1 เมื่อ *K*=2 (ผลการศึกษาแสดงในตารางที่ 4)

- เมื่อประชากรที่ 1 มีการแจกแจงปรกติ และประชากรที่ 2 มีการแจกแจงเอกรูป การทดสอบทีมีกำลังการทดสอบสูงกว่าการทดสอบวิลค็อกซันในทุกกรณี

- เมื่อประชากรที่ 1 และประชากรที่ 2 มีการแจกแจงปรกติ – เบ้ซ้าย การแจกแจงปรกติ – เบ้ขวา การแจกแจงเอกรูป - เบ้ซ้าย และการแจกแจงเอกรูป – เบ้ขวา ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 การทดสอบทีมีกำลังการทดสอบสูงกว่าการทดสอบวิลค็อกซัน เมื่อขนาดตัวอย่างเท่ากับ 10, 20, 30 และ 50 และเมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน อย่างไรก็ตามในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.7 การทดสอบทีมีกำลังการทดสอบสูงกว่า

การทดสอบวิลค็อกซันเมื่อขนาดตัวอย่างเท่ากับ 10, 20 และ 30 และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน

- เมื่อประชากรที่ 1 มีการแจกแจงเบ้ซ้าย และประชากรที่ 2 มีการแจกแจงเบ้ขวา เมื่อขนาดตัวอย่างเท่ากับ 10 การทดสอบทีมีประสิทธิภาพดีกว่าการทดสอบวิลค็อกซัน และเมื่อขนาดตัวอย่างเท่ากับ 20, 30, 50 และ 100 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากันในทุกกรณี

- กำลังการทดสอบของการทดสอบทั้งสองวิธีจะเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น และสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้น

ตารางที่ 4 กำลังการทดสอบของการทดสอบทีและการทดสอบวิลค็อกซัน

การแจกแจง		ขนาด	สัมประสิทธิ์สหสัมพันธ์			
			0.5		0.7	
			การทดสอบ	การทดสอบ	การทดสอบ	การทดสอบ
1	2	ตัวอย่าง	ที	วิลค็อกซัน	ที	วิลค็อกซัน
ปรกติ	เอกรูป	10	0.0368*	-	0.0338*	-
		20	0.0452*	0.0380	0.0492*	0.0394
		30	0.0460*	0.0434	0.0458*	0.0400
		50	0.0502*	0.0496	0.0518*	0.0516
		100	0.0482*	0.0464	0.0566*	0.0534
ปรกติ	เบ้ซ้าย	10	0.4008*	-	0.6126*	-
		20	0.7560*	0.7236	0.9426*	0.9284
		30	0.9074*	0.9016	0.9896*	0.9884
		50	0.9910*	0.9882	0.9996*	0.9996*
		100	1.0000*	1.0000*	1.0000*	1.0000*
ปรกติ	เบ้ขวา	10	0.3860*	-	0.6040*	-
		20	0.7506*	0.7256	0.9386*	0.9214
		30	0.9186*	0.9078	0.9918*	0.9902
		50	0.9892*	0.9886	1.0000*	1.0000*
		100	1.0000*	1.0000*	1.0000*	1.0000*

* การทดสอบที่มีกำลังการทดสอบสูงสุด

ตารางที่ 4 (ต่อ) กำลังการทดสอบของการทดสอบทีและการทดสอบวิลค็อกซัน

การแจกแจง		ขนาด	สัมประสิทธิ์สหสัมพันธ์			
			0.5		0.7	
			การทดสอบ	การทดสอบ	การทดสอบ	การทดสอบ
1	2	ตัวอย่าง	ที	วิลค็อกซัน	ที	วิลค็อกซัน
เอกรูป	เบ้ซ้าย	10	0.3554*	-	0.5398*	-
		20	0.6938*	0.6656	0.8970*	0.8806
		30	0.8726*	0.8644	0.9826*	0.9792
		50	0.9812*	0.9796	0.9996*	0.9996*
		100	1.0000*	1.0000*	1.0000*	1.0000*
เอกรูป	เบ้ขวา	10	0.3432*	-	0.5322*	-
		20	0.7170*	0.6814	0.8936*	0.8724
		30	0.8722*	0.8660	0.9790*	0.9734
		50	0.9844*	0.9824	0.9996*	0.9996*
		100	0.9998*	0.9998*	1.0000*	1.0000*
เบ้ซ้าย	เบ้ขวา	10	0.9520*	-	0.9992*	-
		20	0.9996*	0.9996*	1.0000*	1.0000*
		30	1.0000*	1.0000*	1.0000*	1.0000*
		50	1.0000*	1.0000*	1.0000*	1.0000*
		100	1.0000*	1.0000*	1.0000*	1.0000*

* การทดสอบที่มีกำลังการทดสอบสูงสุด

2.2. เมื่อ $\kappa = 3$ และประชากร 3 กลุ่ม มีการแจกแจงที่แตกต่างกัน 2 ชนิด (ผลการศึกษาแสดงในตารางที่ 5)

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงปรกติ – ปรกติ – เอกรูป และการแจกแจงเอกรูป – เอกรูป – ปรกติ การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman ในทุกกรณี

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงปรกติ – ปรกติ – เบ้ซ้าย การแจกแจงปรกติ – ปรกติ – เบ้ขวา การแจกแจงเอกรูป – เอกรูป – เบ้ขวา การแจกแจงเบ้ซ้าย – เบ้ซ้าย – ปรกติ การแจกแจงเบ้ขวา – เบ้ขวา – ปรกติ และการแจกแจงเบ้ขวา – เบ้ขวา – เอกรูป ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 และเมื่อขนาดตัวอย่างเท่ากับ 100 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.7 การทดสอบเอฟมีค่ากำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 และเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงเอกรูป – เอกรูป – เบ้ซ้าย การแจกแจงเบ้ซ้าย – เบ้ซ้าย – เอกรูป การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 และการทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากันเมื่อขนาดตัวอย่างเท่ากับ 100

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงเบ้ซ้าย – เบ้ซ้าย – เบ้ขวา และการแจกแจงเบ้ขวา – เบ้ขวา – เบ้ซ้าย พบว่าการทดสอบทั้งสองวิธีมีกำลังการทดสอบพอ ๆ กัน และเท่ากับ 1.0000

- กำลังการทดสอบของการทดสอบทั้งสองวิธีมีค่าเพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น และสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้น

2.3. เมื่อ $K = 3$ และประชากร 3 กลุ่ม มีการแจกแจงที่แตกต่างกัน 3 ชนิด (ผลการศึกษาแสดงในตารางที่ 6)

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงปรกติ – เอกรูป – เบ้ซ้าย และการแจกแจงปรกติ – เอกรูป – เบ้ขวา การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20, 30 และ 50 และการทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากันเมื่อขนาดตัวอย่างเท่ากับ 100

- เมื่อประชากรที่ 1, 2 และ 3 มีการแจกแจงปรกติ – เบ้ซ้าย – เบ้ขวา ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20 และ 30 และการทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากันเมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.7 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน

- เมื่อประชากรมีที่ 1, 2 และ 3 มีการแจกแจงเอกรูป – เบ้ซ้าย – เบ้ขวา ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5 การทดสอบเอฟมีกำลังการทดสอบสูงกว่าการทดสอบ Friedman เมื่อขนาดตัวอย่างเท่ากับ 20 และการทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากันเมื่อขนาดตัวอย่างเท่ากับ 30, 50 และ 100 ในกรณีที่สัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.7 การทดสอบทั้งสองวิธีมีกำลังการทดสอบเท่ากัน

ตารางที่ 5 กำลังการทดสอบของการทดสอบเอฟและการทดสอบ Friedman เมื่อประชากร 3 กลุ่มมีการแจกแจงที่ต่างกัน 2 ชนิด

การแจกแจง			ขนาด	สัมประสิทธิ์สหสัมพันธ์			
				0.5		0.7	
				การทดสอบ	การทดสอบ	การทดสอบ	การทดสอบ
1	2	3	ตัวอย่าง	เอฟ	Friedman	เอฟ	Friedman
ปรกติ	ปรกติ	เอกรูป	20	0.0460*	0.0394	0.0494*	0.0448
			30	0.0504*	0.0490	0.0492*	0.0472
			50	0.0488*	0.0478	0.0444	0.0494*
			100	0.0550*	0.0514	0.0506*	0.0498
ปรกติ	ปรกติ	เบ้ซ้าย	20	0.7934*	0.7360	0.9626*	0.9448
			30	0.9392*	0.9034	0.9974*	0.9950
			50	0.9962*	0.9932	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
ปรกติ	ปรกติ	เบ้ขวา	20	0.7964*	0.7344	0.9656*	0.9476
			30	0.9470*	0.9174	0.9976*	0.9940
			50	0.9968*	0.9930	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
เอกรูป	เอกรูป	ปรกติ	20	0.0534*	0.0516	0.0554*	0.0480
			30	0.0502*	0.0452	0.0530*	0.0458
			50	0.0490*	0.0486	0.0514*	0.0488
			100	0.0536*	0.0492	0.0510*	0.0508
เอกรูป	เอกรูป	เบ้ซ้าย	20	0.7228*	0.6368	0.9376*	0.9024
			30	0.9056*	0.8460	0.9908*	0.9820
			50	0.9920*	0.9786	1.0000*	0.9998
			100	1.0000*	1.0000*	1.0000*	1.0000*
เอกรูป	เอกรูป	เบ้ขวา	20	0.7486*	0.6586	0.9348*	0.8984
			30	0.9104*	0.8478	0.9926*	0.9820
			50	0.9924*	0.9792	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*

ตารางที่ 5 (ต่อ) กำลังการทดสอบของการทดสอบเอฟและการทดสอบ Friedman เมื่อประชากร 3 กลุ่มมีการแจกแจงที่ต่างกัน 2 ชนิด

การแจกแจง			ขนาดตัวอย่าง	สัมประสิทธิ์สหสัมพันธ์ 0.5 การทดสอบเอฟ	สัมประสิทธิ์สหสัมพันธ์ 0.5 การทดสอบ Friedman	สัมประสิทธิ์สหสัมพันธ์ 0.7 การทดสอบเอฟ	สัมประสิทธิ์สหสัมพันธ์ 0.7 การทดสอบ Friedman
1	2	3					
เบ้ซ้าย	เบ้ซ้าย	ปรกติ	20	0.8050*	0.7412	0.9614*	0.9408
			30	0.9358*	0.8974	0.9966*	0.9932
			50	0.9972*	0.9926	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
เบ้ซ้าย	เบ้ซ้าย	เอกรูป	20	0.7630*	0.6476	0.9448*	0.8968
			30	0.9232*	0.8456	0.9932*	0.9778
			50	0.9940*	0.9726	0.9994*	0.9990
			100	1.0000*	1.0000*	1.0000*	1.0000*
เบ้ซ้าย	เบ้ซ้าย	เบ้ขวา	20	1.0000*	1.0000*	1.0000*	1.0000*
			30	1.0000*	1.0000*	1.0000*	1.0000*
			50	1.0000*	1.0000*	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
เบ้ขวา	เบ้ขวา	ปรกติ	20	0.8036*	0.7350	0.9638*	0.9448
			30	0.9492*	0.9164	0.9978*	0.9954
			50	0.9966*	0.9898	1.0000*	1.0000*
			100	1.0000*	0.9998	1.0000*	1.0000*
เบ้ขวา	เบ้ขวา	เอกรูป	20	0.7600*	0.6458	0.9366*	0.8908
			30	0.9162*	0.8412	0.9922*	0.9782
			50	0.9908*	0.9730	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
เบ้ขวา	เบ้ขวา	เบ้ซ้าย	20	1.0000*	0.9990	1.0000*	1.0000*
			30	1.0000*	1.0000*	1.0000*	1.0000*
			50	1.0000*	1.0000*	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*

* การทดสอบที่มีกำลังการทดสอบสูงสุด

ตารางที่ 6 กำลังการทดสอบของการทดสอบเอฟและการทดสอบ Friedman เมื่อประชากร 3 กลุ่ม มีการแจกแจงที่ต่างกัน 3 ชนิด

การแจกแจง			ขนาดตัวอย่าง	สัมประสิทธิ์สหสัมพันธ์ 0.5 การทดสอบเอฟ	0.5 การทดสอบ Friedman	0.7 การทดสอบเอฟ	0.7 การทดสอบ Friedman
1	2	3					
ปรกติ	เอกรูป	เบ้ซ้าย	20	0.7608*	0.6860	0.9418*	0.9154
			30	0.9310*	0.8890	0.9940*	0.9896
			50	0.9958*	0.9894	1.0000*	0.9998
			100	1.0000*	0.9998	1.0000*	1.0000*
ปรกติ	เอกรูป	เบ้ขวา	20	0.7638*	0.6870	0.9528*	0.9264
			30	0.9246*	0.8846	0.9944*	0.9898
			50	0.9946*	0.9878	1.0000*	0.9998
			100	1.0000*	1.0000*	1.0000*	1.0000*
ปรกติ	เบ้ซ้าย	เบ้ขวา	20	0.9996*	0.9980	1.0000*	1.0000*
			30	1.0000*	0.9998	1.0000*	1.0000*
			50	1.0000*	1.0000*	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*
เอกรูป	เบ้ซ้าย	เบ้ขวา	20	0.9994*	0.9980	1.0000*	1.0000*
			30	1.0000*	1.0000*	1.0000*	1.0000*
			50	1.0000*	1.0000*	1.0000*	1.0000*
			100	1.0000*	1.0000*	1.0000*	1.0000*

* การทดสอบที่มีกำลังการทดสอบสูงสุด

สรุปผลการวิจัย

จากผลการวิจัยจะเห็นได้ว่าสำหรับข้อมูลที่อยู่ในมาตราอันดับแบบมาตรวัดของลิเกิร์ต 5 ระดับ ไม่ว่าการแจกแจงความน่าจะเป็นของข้อมูลจะเป็นชนิดใด การทดสอบทีและเอฟที่ใช้สำหรับทดสอบค่ากลางของประชากร 2 และ 3 กลุ่ม ที่ไม่เป็นอิสระกัน ซึ่งเป็นการทดสอบอิงพารามิเตอร์มีประสิทธิภาพดีกว่าการทดสอบวิลค็อกซันและการทดสอบ Friedman ซึ่งเป็นการทดสอบไม่อิงพารามิเตอร์ ทั้งในกรณีของความสามารถในการควบคุมความน่าจะเป็นของความผิดพลาดแบบที่ 1 และกำลังการทดสอบ ทั้ง ๆ ที่การทดสอบทีและเอฟมีข้อสมมุติว่าข้อมูลที่ศึกษาต้องมาจากประชากรที่มีการแจกแจงปรกติ และอยู่ในมาตราช่วงเป็นอย่างน้อย ซึ่งผลการวิจัยนี้สอดคล้องกับศึกษาของ Norman (2010) ที่แสดงให้เห็นว่าการทดสอบอิง

พารามิเตอร์เป็นการทดสอบเชิงสถิติที่มีความแกร่งต่อการละเมิดข้อสมมุติ และสอดคล้องกับการศึกษาของ Romano, Kromrey, Coraggio and Skowronek (2006) ที่กล่าวว่าการทดสอบทีเป็นวิธีที่เหมาะสมในกรณีที่ต้องการทดสอบความแตกต่างระหว่างกลุ่มของข้อมูลที่อยู่ในมาตราอันดับ นอกจากนั้นการสร้างข้อมูลมาตราอันดับในการศึกษานี้มีลักษณะที่ระดับของข้อมูล 1 2 3 4 และ 5 มีช่วงของความแตกต่างในแต่ละระดับที่เท่ากัน ทำให้พิจารณาได้อีกนัยหนึ่งว่าข้อมูลลักษณะนี้เป็นข้อมูลที่อยู่ในมาตราช่วง และมีคุณสมบัติตรงตามข้อสมมุติของสถิติทดสอบทีและเอฟ ซึ่งหากข้อมูลที่เก็บรวบรวมได้จากการศึกษาจริงไม่ได้มีลักษณะดังข้อมูลที่สร้างขึ้นในการศึกษานี้ แต่เป็นข้อมูลที่อยู่ในมาตราอันดับอย่างแท้จริง นั่นคือ ช่วงของความแตกต่างในแต่ละระดับอาจจะไม่เท่ากัน การทดสอบทีและเอฟอาจจะไม่มีประสิทธิภาพดีที่สุดก็เป็นได้ ดังนั้น ในการเลือกใช้การทดสอบทางสถิติที่เหมาะสมผู้วิจัยยังควรต้องคำนึงถึงข้อสมมุติของสถิติทดสอบนั้น ๆ อยู่

กิตติกรรมประกาศ

ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่สนับสนุนทุนอุดหนุนการวิจัยประเภทเงินรายได้ ประจำปีงบประมาณ 2559

เอกสารอ้างอิง

Barbiero, A. & Ferrari, P. A. (2015). *GenOrd: Simulation of ordinal and discrete variables with given correlation matrix and marginal distributions*. Retrieved January 10, 2016 from http://CRAN.R project.org/package =GenOrd.

Bradley, J. V. (1978). Robustness?, *British Journal of Mathematical and Statistical Psychology, 31*(1), 144-152.

Ferrari, P. A. & Barbiero, A. (2012). Simulation ordinal data. *Multivariate Behavioral Research*, *47*(4), 566-589.

Harwell, M. R. & Serlin, R. C. (1994). A Monte Carlo study of the Friedman test and some competitors in the single factor, repeated measures design with unequal covariances. *Computational Statistics and Data Analysis*, *17*(1), 35-49.

Norman, G. (2010). Likert scales, levels of measurement and the “Laws” of statistics. *Advance in Health Science Education, 15*(5), 625-632.

Romano, J., Kromrey, J.D., Coraggio, J., & Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen’s d for evaluating group differences on the NSSE and other surveys?. Paper presented at *the annual meeting of the Florida Association of Institutional Research, February 1 -3, 2006,* Cocoa Beach, Florida.

Zimmerman, D.W. & Zumbo, B.D. (1993). Relative power of the Wilcoxon test, the Friedman test, and repeated measures ANOVA on ranks. *The Journal of Experimental Education, 62*(1), 75-86.